

```
In [12]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
import scipy.stats as stats
```

```
In [13]: mon = pd.read_csv("Monday.csv", index_col=[0])
```

```
In [14]: tue = pd.read_csv("Tuesday.csv", index_col=[0])
```

```
In [15]: wed = pd.read_csv("Wednesday.csv", index_col=[0])
```

```
In [16]: thur = pd.read_csv("Thursday.csv", index_col=[0])
```

```
In [17]: fri = pd.read_csv("Friday.csv", index_col=[0])
```

```
In [18]: arr = ("mon", "tue", "wed", "thu", "fri")
count = 0
for i in (mon, tue, wed, thur, fri):
    i["timestamp"] = pd.to_datetime(i["index"])
    i.set_index("timestamp", inplace=True)
    i.drop(columns={"index"}, inplace=True)
    i[arr[count]] = [1 for i in range(0, len(i))]
    i.head()
    count = count + 1

#     thur["timestamp"] = pd.to_datetime(thur["index"])
#     thur.set_index("timestamp", inplace=True)
#     thur["count"] = [1 for i in range(0, len(thur))]
#     thur.head()
```

```
In [19]: j = pd.DataFrame()
for i in (mon, tue, wed, thur, fri):
    resampled = i.resample("20Min").sum()
    resampled.index = pd.Series(resampled.index).apply(lambda x : x.time())
    j = pd.concat([j, resampled], axis=1)
```

```
In [20]: c = j.dropna()
```

```
In [21]: c.mean()
```

```
Out[21]: mon    2.307692
tue    2.538462
wed    2.923077
thu    2.230769
fri    2.538462
dtype: float64
```

```
In [22]: j.mean()
```

```
Out[22]: mon    2.222222
         tue    2.400000
         wed    2.812500
         thu    2.565217
         fri    2.142857
         dtype: float64
```

```
In [23]: j.count()
```

```
Out[23]: mon    18
         tue    15
         wed    16
         thu    23
         fri    28
         dtype: int64
```

```
In [24]: weighted_average = (j.mean()*j.count()).sum()/j.count().sum()
```

```
In [25]: round(weighted_average,2)
```

```
Out[25]: 2.4
```

Mean frequency in 2 minute intervals for arrival of blue buses is 0.26 buses; or we can expect 0.13 buses a minute

Conversely, we need to wait an average of 7.69 minutes for the blue bus

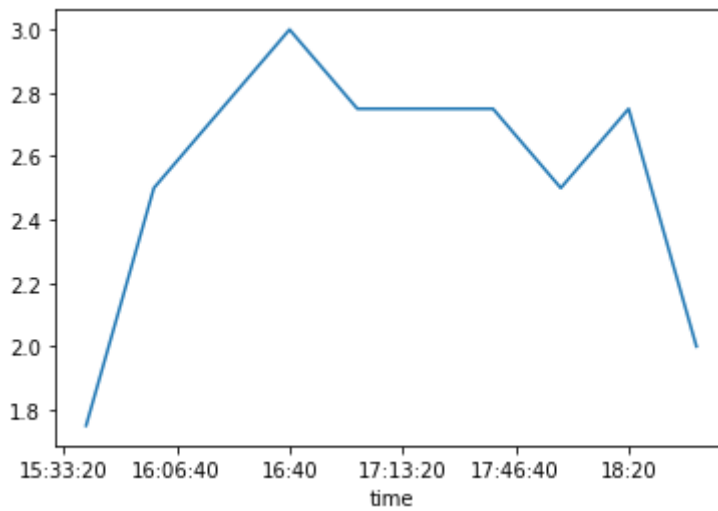
```
In [26]: import math
```

```
In [81]: roll = pd.DataFrame()
         for i in ("mon", "tue", "wed", "thu", "fri"):
             roll = pd.concat([roll, c[i].rolling(4).mean()[3:]], axis=1)

         roll = roll.rename(columns={'mon': 'Monday', 'tue': 'Tuesday', 'wed': 'Wednesday', 't
```

```
In [92]: roll["Monday"].plot()
```

```
Out[92]: <AxesSubplot:xlabel='time'>
```



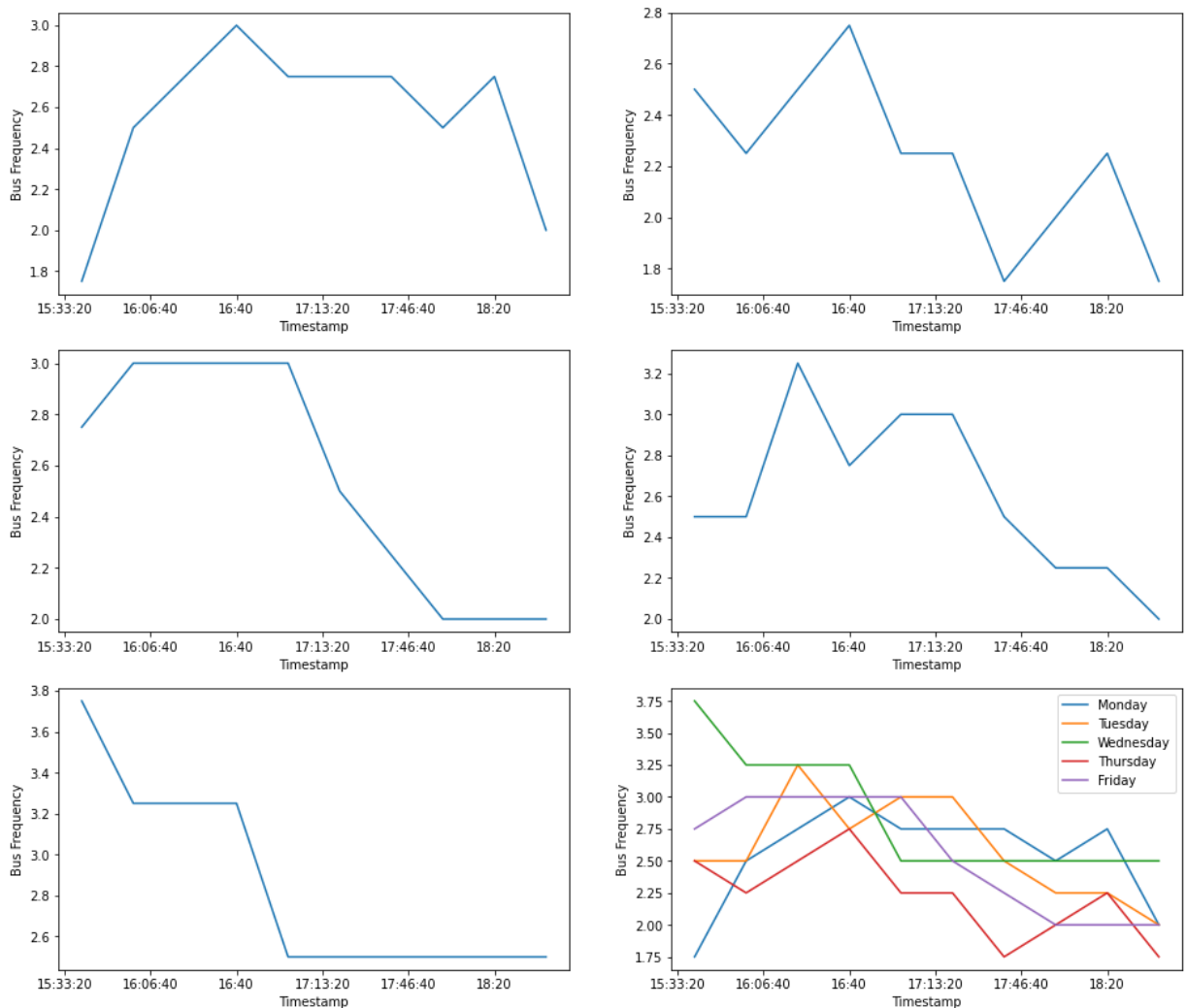
In [103...

```
fig, axis = plt.subplots(3,2, figsize=(16,14))
counter=0

for i in ('Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday'):
    roll[str(i)].plot(ax=axis[counter%3][counter%2], xlabel = "Timestamp", ylabel = "
    counter=counter+1
    roll.plot(ax=axis[2][1], xlabel = "Timestamp", ylabel = "Bus Frequency")
```

Out[103...

<AxesSubplot:xlabel='Timestamp', ylabel='Bus Frequency'>



We observe the same general trend, most buses are in

the interval from 4pm-5pm, after which their numbers steadily decline into the evening

In [568...

```
j = pd.DataFrame()
for i in (mon,tue,wed,thur,fri):
    resampled=i.resample("5Min").sum()
    resampled.index=pd.Series(resampled.index).apply(lambda x : x.time())
    j = pd.concat([j,resampled],axis=1)
c = j.dropna()
```

In [578...

```
j = pd.DataFrame()
for i in (mon,tue,wed,thur,fri):
    resampled=i.resample("2Min").sum()
    resampled.index=pd.Series(resampled.index).apply(lambda x : x.time())
    j = pd.concat([j,resampled],axis=1)
fig, axis = plt.subplots(5,1, figsize=(10,20))
counter=0
for i in ("mon","tue","wed","thu","fri"):
    sb.histplot(data = j[i], ax=axis[counter])
    counter=counter+1
```

